PATENT APPLICATION
Navy Case No. **84,655**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

TO ALL WHOM IT MAY CONCERN:

BE IT KNOWN THAT **Jonathan Schuler, Dean Scribner and J. Grant Howard** who are citizens of the United States of America, and are residents of Fairfax, VA, Arlington, VA and Woodbridge, VA, invented certain new and useful improvements in **"AN ALGORITHMIC TECHNIQUE FOR INCREASING THE SPACIAL ACUITY OF A FOCAL PLANE ARRAY ELECTRO-OPTIC IMAGING SYSTEM"** of which the following is a specification:

RECEIVED
MAR 25 2004
OIPE/JCWS

Please Contact Preparer:
John Gladstone Mills III
Reg. No. 49,054
Tel: 202-404-1553
Date: March 11, 2004

AN ALGORITHMIC TECHNIQUE FOR INCREASING THE SPACIAL ACUITY OF A FOCAL PLANE ARRAY ELECTRO-OPTIC IMAGING SYSTEM

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention deals generally with an algorithm for increasing the spatial acuity of Focal Plane Array based Electro-Optic imaging systems by accumulating multiple frames of imagery into a single composite image and thus reducing the effective focal length of a viewing lens.

2. Description of the Related Prior Art

This invention relates to particular types of imaging systems, and more specifically, to a method that improves the spatial resolution of such imaging systems. This is achieved by assimilating a video sequence of images that may drift, yet dwell, over an object of interest into a single composite image with higher spatial resolution than any individual frame from the video sequence.

This technique applies to a particular class of non-coherent electro-optical imaging systems that consist of a lens projecting incoming light onto a focal plane. Positioned at the focal plane is an array of electronic photo-conversion detectors, whose relative spatial positions at the focal plane are mechanically constrained to be fixed, such

as through lithography techniques common to the manufacturing processes of focal plane array detectors.

It is noted this invention cannot increase the physically achievable resolution of an imaging system, which is fundamentally bounded by the diffraction limit of a lens with finite aperture at a given wavelength of non-coherent light. Rather, this invention recovers for resolution that is additionally lost to distortions of noise, aliasing, and pixel blur endemic to any focal plane array detector.

In conventional optical sensor design, the lens aperture size determines the diffraction limited resolution, in angle, of an optic at a specific wavelength. The lens projects this resolution limit as a blur-circle, or point-spread function, at the focal plane of the sensor. The actual size of the point spread function at the focal plane is geometrically related, and directly proportional to the focal length of the lens. In order for a focal plane array, with finite sized pixels, to sufficiently sample the projected optical image without alias distortion, the projected point spread function must be sufficiently large to span at least two to three pixels. This loose constraint places a bound on the minimum necessary focal length of a lens to eliminate alias distortion in the imagery captured by a focal plane array. The described invention synthetically increases the pixel density of the focal plane array. Thus, it reduces the necessary size of the projected blur circle, or equivalently, it reduces the minimum focal length required to eliminate alias distortion. The described invention permits optical sensors with a fixed size aperture to

deploy lenses with shorter focal length that are more compact, weight less, and offer wider field of views, while maintaining system acuity.

Single frame digital image restoration is a widely implemented mathematical technique that can compensate for known or estimated distortions endemic to a given digital image, improving the perceptual acuity and operational resolution of the constituent digital imaging sensor. (See Chapter 8 of Fundamentals of Digital Image Processing, A.K. Jain, Prentice Hall 1989)

The performance of such single-frame restoration techniques can be bounded by two limitations:

1) Insufficient spatial sampling of the projected optical image when measured by a single-frame capture of the projected optical image by a focal plane array. Depending on the F-number of the lens and the physical spacing (pixel pitch) of detectors, this situation may result in spatial alias distortion that is unrecoverable in a general sense.
2) Noise of the constituent pixel detectors in a focal plane array, and the associated read-out electronic microcircuits, which will limit the performance of any subsequent restoration filter.

When imaging an object of interest, a sensor may often stare at that object for sufficient time to create a video sequence of images that dwell, with the possibility to drift, over the particular object. For many applications, only a single frame is recorded

and processed, discarding the statistically innovative information that may be contained in additional, but unexamined images captured by the focal plane array.

Straightforward implementation of resolution enhancement through multiple frames of imagery have been implemented by controlled micro-dither scanning of a sensor (W.F. O'Neal "Experimental Performance of a Dither-Scanned InSb Array" Proceedings on the 1993 Meeting of the IRIS Specialty Group on Passive Sensors), where a stationary scene is imaged by a sensor subject to a well controlled pattern of orientation displacements, such as an integer fraction of a pixel. Image recovery is then implemented by appropriately interlacing the constituent images into a composite image with an integer-multiple increase in sampling density. Such techniques are very effective in suppressing alias distortions of any single frame, but may come at the cost of stabilization requirements that limit their implementation in practical, man-portable sensor systems. Without any deliberate dithering, such video sequences of images may still be subject to unknown displacements, which can be exploited to provide the same benefits as controlled dither. There has been a history of research in algorithms to implement a multi-frame image restoration on such data sets (T. S. Huang., "Multiple frame image restoration and registration," in Advances in Computer Vision and Image Processing, vol. 1, JAI Press, 1984.). The preponderance of these algorithms follows a common, non-linear approach to this problem:

1) Pre-suppose the existence of a high-resolution image, perhaps sampled at some integer multiple of the number of pixels of the constituent images. Seed this high

resolution image with some initial guess, such as the interpolation of any single frame to the higher spatial sampling rate.

2) Derive some guess of the motion of the video sequence relative to the high resolution image. Displace and down-sample the high resolution image so as to create a synthetic video sequence consistent with the observed video sequence.
3) Determine some form of error between the synthetic and actual video sequence.
4) Adjust the estimates of both the high-resolution image and the scene motion so as to reduce the error between synthetic and actual video sequences.
5) Repeat steps 3 & 4 until a convergence in error has been reached.

This approach to multi-frame image restoration is plagued by three limitations

1) Iterative algorithms often exhibit long convergence times and are computationally intense.
2) Numerical techniques for adjusting the estimates of step 4 often depend on specifying an underlying probability distribution model. Such Maximum Likelihood or Maximum A-Postori techniques prove to be numerically unstable if the underlying data deviates from such idealized statistical models.
3) Many such algorithms are constrained to cases of simple motion models, such as uniform displacements between frames of video, which may not fully represent the true motion of the sequence.
4) Final restoration of the high resolution image additionally depends on an empirical smoothing kernel with little or no analytic derivation.

SUMMARY OF THE INVENTION

This invention relates to particular types of imaging systems, and more specifically, to a method that improves the spatial resolution of such imaging systems. This is achieved by assimilating a video sequence of images that may drift, yet dwell, over an object of interest into a single composite image with higher spatial resolution than any individual frame from the video sequence.

This technique applies to a particular class of non-coherent electro-optical imaging systems that consist of a lens projecting incoming light onto a focal plane. Positioned at the focal plane is an array of electronic photo-conversion detectors, whose relative spatial positions at the focal plane are mechanically constrained to be fixed, such as through lithography techniques common to the manufacturing processes of focal plane array detectors.

It is noted this invention cannot increase the physically achievable resolution of an imaging system, which is fundamentally bounded by the diffraction limit of a lens with finite aperture at a given wavelength of non-coherent light. Rather, this invention recovers for resolution that is additionally lost to distortions of noise, aliasing, and pixel blur endemic to any focal plane array detector. This process is implemented on a computational platform that acquires frames of digital video from an imaging system that drifts, yet dwells on an object of interest. This process assimilates this video sequence into a single image with improved qualities over that of any individual frame from the

original sequence of video. A restoration process can then be applied to the improved image, resulting in operational image acuity.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a preferred embodiment of the invention in a computer processing system

Figure 2 illustrates a flow chart outlining the operation sequence of the invention

Figure 3 illustrates a sequence of video imagery, along with corresponding coordinate directions. Additionally, the template image is highlighted.

Figure 4 illustrates a sequence of vector field plots, corresponding to the displacement estimated for every pixel of the video sequence illustrated in FIGURE 3.

Figure 5 illustrates, in MATLAB script, the algorithm that implements an estimate of nearest pixel image displacement, by image correlation.

Figure 6 illustrates the correlation surface corresponding to two images of the same scene subject to sensor motion.

Figure 7 illustrates, in MATLAB script, an algorithm that implements sub-pixel image displacement by numerical solution to the Brightness Constancy Constraint (BCC)

equation (Algorithm described in Digital Video Processing, A. M. Tekalp, 1995 Prentice Hall, pp 81-86)

Figure 8 illustrates the coordinate topology of focal plane array (FPA) sensors. In particular, every pixel can be addressed by an ordered pair of whole-integers. Such an address also corresponds to the physical location of a given photo-detector pixel of the FPA.

Figure 9 illustrates a flow chart detailing the process by which a pixel in a high resolution composite image is estimated from pixels of original video.

Figure 10 illustrates the high-resolution lattice data structure associated with the re-sorted image data. Of note is that every lattice site can be variably populated by a differing number of pixels from the video sequence whose estimated coordinates lie within the coordinate span of the high-resolution lattice site.

DETAILED DESCRIPTION OF THE INVENTION

This preferred embodiment of this process is on a digital imaging system illustrated in FIGURE 1. This system consists of a digital imaging sensor or camera, 101, consisting of a lens that focuses light, 100, onto a focal plane array of photo-detectors that produces an electronic representation of the projected optical image of the lens. The data from this camera is then captured by some form of a computing platform, 102, such as a personal computer, laptop, handheld digital assistant, or any processing devices embedded within the camera, 101. Such a computing platform may also store captured image sequences for long durations on non-volatile media, 104. This computing platform is also capable of implementing the described process, rendering an image that is presented to the operator through some display device, 103.

The process of increasing the spatial acuity of Focal Plane Array based Electro-Optic imaging systems by accumulating multiple frames of imagery into a single composite image is illustrated in the process flow chart of FIGURE 2.

The initial pre-processing steps will include the following

1) Launching the software on the processing platform, 201. This may be done automatically with activation of the camera sensor, 202.
2) Collection of a video sequence with suitable motion displacement between frames, 203, or loading in a previously recorded suitable sequence from non-volatile digital data storage, 204.

3) From the acquired or loaded video sequence, select a subset of video frame which will be integrated into a final composite image, 205.
4) From the selected subset of video frames, select one particular frame which will serve as the template frame for subsequent restoration, 206.
5) From the template frame, select a particular spatial Region of Interest (ROI) that will be restored, 207.
6) Additionally, select a factor by which the spatial sampling of the digital image will be restored, 208.

Given such a configuration, multi-frame image restoration can be achieved in three further stages of processing illustrated in the process flow chart of FIGURE 2.

1) *Motion Estimation of a video sequence*, 209.
2) *Assembly of video frames into a single composite image based on estimated positions of individual pixels*, 210.
3) *Restoration of the composite image,* 211

These three stages of processing are further elaborated as follows:

Step 1: Motion Estimation of a video sequence.

The motion of pixels in a video sequence can be characterized by the optical flow, defined as a mapping that relates the spatial coordinates of all pixels of a video sequence. Mathematically, the optical flow estimation problem is ill posed (referred as the "aperture problem"), and requires additional regularization constraints to generate a solution for

this mapping between the spatial coordinates of pixels. Such regularization introduces a bias-variance trade in the motion estimation, between bias against sensitivity to spatially localized motion versus an increase in overall statistical variance of the motion estimator.

In this embodiment, a single image, 302, from the sequence, 301-304, is taken to serve as a template, as shown in FIGURE 3. The motion of all other frames of video is estimated relative to this template image, as shown in FIGURE 4. The motion of any particular frame can be described by a corresponding tensor field, 401-404, where every 2 dimensional pixel coordinate has associated a 2 dimensional vector corresponding to the pixel displacement relative to the corresponding pixel coordinate of the template frame. Because there is no motion of the template image with respect to itself, its corresponding motion field, 402, will be trivial arrays of zeros. In the current embodiment, the motion is assumed to be a uniform displacement. This uniform displacement is estimated by a two-stage procedure:

1) Estimate nearest-pixel displacement by image correlation. This is illustrated in by the MATLAB code of FIGURE 5, as well as the correlation surface for two sequential frames of digital video, 601-602, illustrated in FIGURE 6, where the location of the peak, 604, of this correlation surface, 603, corresponds to the displacement between the two images.
2) Given this estimate of nearest pixel displacement, re-crop each image accordingly so that the cropped images have the same size and are pixel aligned. Then, estimate the sub-pixel displacement between the cropped images by a least-

squares solution to a brightness-constancy-constraint (BCC) model of the video sequence, which is illustrated in the MATLAB code of FIGURE 7.

Whereas every pixel in the template image is tagged with a whole-integer coordinate consistent with the address coordinate of the corresponding focal plane array detector, as shown in FIGURE 8. Pixels in every other frame are tagged with an adjusted coordinate based on the displacement estimate of their frame. From these tagged coordinates, a high resolution composite image can be assembled from individual pixels across different low resolution frames of constituent video. Extensions to this embodiment can include more complicated motion models relating coordinates between frames of video, such as affine, bilinear, or polynomial model distortions to accommodate perspective changes or geometric lens distortions. Additionally, any estimators used to determine the motion displacement between frames of video are themselves statistical operations with intrinsic uncertainty. Further extensions to this embodiment can include some additional estimate of the statistical uncertainty, such as a confidence interval, associated with each estimated coordinate for every pixel.

After motion estimation has been applied to a video sequence, every pixel of the video sequence will have associated 5 quantities relevant to subsequent image restoration of the ROI of the template frame, namely:

1) Pixel intensity
2) X-coordinate location
3) Y-coordinate location

4) X-coordinate estimate uncertainty

5) Y-coordinate estimate uncertainty

Step 2: Assembly of video frames into a single composite image based on estimated positions of individual pixels

Motion estimation, applied to a video sequence, generates a 5-entity database for every pixel element consisting of:

1) Pixel intensity
2) X-coordinate location
3) Y-coordinate location
4) X-coordinate location uncertainty
5) Y-coordinate location uncertainty

This database of information is then re-assembled into a single composite image according to the following process, as illustrated in FIGURE 9.

1) Define and construct a lattice with a higher sampling density than the template image used for motion estimation, 901. This lattice array does not necessarily have to be a whole number multiple of the template image pixel array size. In certain applications, it may be desirable to make the lattice size the same as the template image size.
2) Compute for each lattice site an associated coordinate interval, 902, corresponding to the rectangular span of each lattice site relative to the template

image coordinate grid, 801. Such coordinate intervals of the lattice image may well span a sub-pixel sized area of the template image coordinates.

3) Find and select all pixels whose coordinates fall within the rectangular span of each lattice site, 903.
    a. In refined implementations of this method that include confidence intervals associated with each pixel's estimated coordinate, one would instead seek all pixels in the database whose estimated coordinates most likely fall within the rectangular span of each lattice site.
4) Given the sample of pixels intensities selected through step 2, construct an aggregate estimator for the estimated pixel intensity of each lattice site, 903. Such an aggregate estimator that can include, but is not limited to, the sample mean, sample median, or any other statistical estimator.
    a. In refined implementation, additional techniques including, but not limited to, statistical bootstrapping can be applied to provide an estimate of the statistical variability associated with estimated intensity of each lattice site, 904.
    b. In refined implementation, additional techniques may adopt kernelling methods estimate intensity using both the data binned in a particular lattice site, as well as the data binned in some region of neighboring lattice sites.

There can be considerable variability in the computational time needed to sort video pixels, 1001, into their appropriate lattice site, 1002, depending on the

implementation of a sorting procedure and computational hardware. A "divide-and-conquer" approach, where the collection of pixels is separated into disjoint collections based on coarse pixel location, will speed up computational time by reducing the number of database elements each lattice site must finally sort through. The particular level of decimation, as well as any recursive implementation of this approach, will depend on the number of available processors, thread messaging speeds, and memory bus access times of the computational hardware upon which this process is implemented.

Step 3: Restoration of the composite image

Once a composite image has been reconstructed in step 2 from the motion estimate information computed in step 1, one can apply any of a myriad of single-frame image restoration techniques, 905, such as Wiener Filtering (Fundamentals of Digital Image Processing, A.K. Jain, Prentice Hall 1989), Lucy–Richardson blind-deconvolution (1972, J. Opt. Soc. Am., 62,55), Pixon-based deconvolution (1996, Astronomy and Astrophysics, 17,5), or other techniques. In refined implementations of this technique, the estimated uncertainty associated with each pixel's intensity of the reconstructed lattice can be leveraged by many single-frame image restoration algorithms to further enhance acuity performance of the restoration.

The performance of any single-frame image restoration technique will invariably improve when applied instead to the composite image derived from multiple frames of video, in so far that the composite image will exhibit:

1) Reduced, or completely eliminated alias distortion resulting from under-sampling of the projected optical image by the FPA detector in the imaging sensor. Such alias distortion would otherwise limit the performance of single frame restoration algorithms applied to only a single frame from a video sequence.
2) Reduced noise associated with the use of aggregate statistical estimators to determine every pixel's intensity in the composite image based on the sub-sample of video pixels estimated to fall within the coordinates of a composite pixel's spatial location. The performance of single-frame image restoration algorithms improves as the constituent noise of the un-restored image is reduced.
3) Empirical estimates of the noise associated with every composite image pixel's estimated intensity, derived from the sub-sample of video pixels estimated to fall within the coordinates of a composite pixel's spatial location. Many single-frame restoration algorithms make an assumption of the underlying noise properties of the un-restored image. Empirical measurements of this underlying noise will improve the accuracy of the underlying assumptions, and consequently improve the performance of the single-frame image restoration step.

Although this invention has been described in relation to an exemplary embodiment thereof, it will be understood by those skilled in the art that still other variations and modifications can be effected in the preferred embodiment without detracting from the scope and spirit of the invention as described in the claims.